

# Projet Base de Données NoSQL

Réalisation d'un tableau de bord à l'aide de MongoDB charts

## Performance des compagnies aériennes aux Etats-Unis

Document rédigé par :

Oumar KANE

Professeur :

Mr PREIRA

AIRLINE DASHBOARD

By

Oumar Kane

**AIRLINE DASHBOARD**

By  
Oumar Kane

## Table des matières

INTRODUCTION .....	5
I. DEFINITION .....	6
1. MongoDB .....	6
2. MongoDB Atlas .....	6
3. MongoDB Charts .....	7
4. MongoDB Compass .....	7
5. Cluster .....	7
II. PROCEDURE D'INSTALLATION .....	8
III. DESCRIPTION DU JEU DE DONNEES .....	22
1. Du point de vue Géographique.....	24
a. Retards par Etat .....	24
b. Nombres d'annulations et de déviations par Etat .....	28
2. Du point de vue Temporel.....	34
V. TABLEAU DE BORD .....	38

**AIRLINE DASHBOARD**

By  
Oumar Kane

# INTRODUCTION

Ce document a pour objectif de présenter les différentes étapes de réalisation d'un tableau de bord sur MongoDB charts avec le jeu de données *Airlines Delay*. Le but est d'offrir au lecteur une bonne compréhension du projet et de lui permettre de réaliser plus facilement des tableaux de bord à partir des données de son cloud à l'aide de **MongoDB Charts**.

Le document comporte 5 parties :

- Définition des termes principaux ;
- Description de la procédure d'installation de MongoDB Charts à l'aide d'illustrations ;
- Description du jeu de données *Airlines Delay* ;
- Mise en œuvre des graphiques ;
- Présentation du tableau de bord final.

# I. DEFINITION

## 1. MongoDB

Apparue au milieu des années 2000, MongoDB est une base de données NoSQL orientée document développée par la société 10gen. Elle est utilisée pour le stockage de volumes massifs de données.

Contrairement à une base de données relationnelle SQL traditionnelle, MongoDB ne repose pas sur des tableaux et des colonnes. Les données sont stockées sous forme de collections et de documents.

## 2. MongoDB Atlas

MongoDB Atlas est la première base de données cloud<sup>1</sup> à permettre aux clients d'exécuter des applications simultanément sur tous les principaux fournisseurs cloud.

Basé sur le serveur<sup>2</sup> NoSQL open source MongoDB, il permet aux clients de déployer simultanément une base de données entièrement gérée et distribuée sur Amazon Web Services (AWS), Google Cloud et Microsoft Azure. Cela signifie que les clients peuvent profiter des avantages du déploiement d'applications sur plusieurs fournisseurs de cloud sans la complexité opérationnelle supplémentaire de la gestion de la réplication et de la migration des données entre les cloud.

Aujourd'hui, les entreprises sont limitées à l'exécution d'applications individuelles sur un seul fournisseur de cloud. Les clusters multicloud de MongoDB Atlas donnent la possibilité aux organisations d'étendre la portée géographique de leurs applications en leur permettant de répliquer les données des trois principaux fournisseurs de cloud vers l'une des 79 régions cloud actuellement prises en

---

<sup>1</sup> Le terme « cloud » désigne les serveurs accessibles sur Internet, ainsi que les logiciels et bases de données qui fonctionnent sur ces serveurs. (source : [Qu'est-ce que le cloud? | Définition du cloud | Cloudflare](#))

<sup>2</sup> Le terme **serveur** désigne le rôle joué par un appareil matériel destiné à offrir des services à des clients en réseau Internet ou intranet. (source : [Serveur \(informatique\) : définition, traduction \(journaldunet.fr\)](#))

charge dans le monde. Les clusters multicloud offrent également aux entreprises une flexibilité sans précédent pour migrer de manière transparente leurs données, un défi généralement onéreux, d'un fournisseur de cloud à un autre afin de répondre aux besoins changeants de leur application ou de leurs exigences métier.

### 3. MongoDB Charts

MongoDB Charts est le moyen le plus simple et le plus rapide de créer des visualisations de données MongoDB. Il est possible de créer des graphiques en quelques minutes, puis d'assembler et de partager des tableaux de bord<sup>3</sup>. Avec les SDK<sup>4</sup> de charts, il devient très facile d'intégrer des graphiques dans des application, puis de les enrichir avec des visualisations de données dynamiques.

### 4. MongoDB Compass

Compass est un outil interactif qui permet d'interroger, d'optimiser et d'analyser des données MongoDB. Il est possible de se connecter à un cluster de MongoDB Atlas à partir de Compass et d'y importer des données en local.

### 5. Cluster

Au sein d'un système informatique, un cluster de serveurs est un groupe de serveurs et d'autres ressources indépendantes fonctionnant comme un seul système. Les serveurs sont généralement situés à proximité les uns des autres, et sont interconnectés par un réseau dédié. Ainsi, les clusters permettent de profiter d'une ressource de traitement de données centralisée. Un client dialogue avec le groupe de serveurs comme s'il s'agissait d'une seule machine.

---

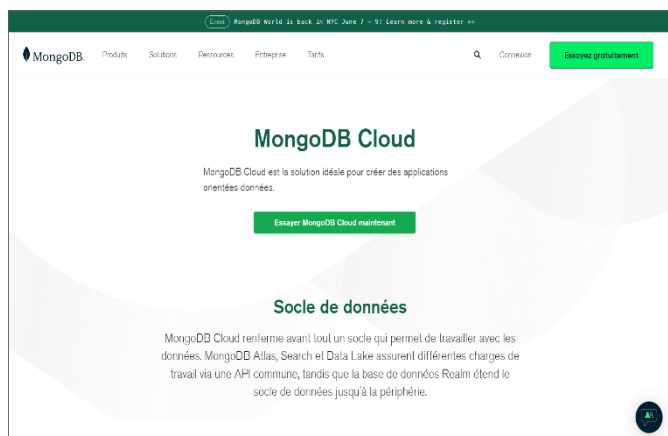
<sup>3</sup> Le **tableau de bord est un outil de pilotage** d'une entreprise constitué de plusieurs indicateurs de performance et qui a pour principales vocations d'anticiper les évolutions prévisibles et d'inciter le chef d'entreprise à prendre des décisions. (source : [Le tableau de bord, outil de pilotage de l'entreprise \(lecoindesentrepreneurs.fr\)](https://lecoindesentrepreneurs.fr/))

<sup>4</sup> Un kit de développement logiciel (SDK) est un ensemble d'outils fourni avec une plateforme matérielle (généralement), un système d'exploitation ou un langage de programmation. (source : [Un SDK, qu'est-ce que c'est ? \(redhat.com\)](https://redhat.com/))

## II. PROCEDURE D'INSTALLATION

Pour pouvoir accéder à MongoDB Charts et créer des graphiques interactifs sur un jeu de données il faut disposer d'un compte MongoDB et créer un cluster à partir de MongoDB Atlas. Le jeu de données à utiliser est également à importer au niveau d'un cluster à l'aide du logiciel MongoDB Compass dont le setup d'installation est téléchargeable à l'adresse suivante [MongoDB Compass | MongoDB](#).

Les illustrations qui suivent comportent les étapes essentielles à suivre pour commencer à utiliser MongoDB Charts :



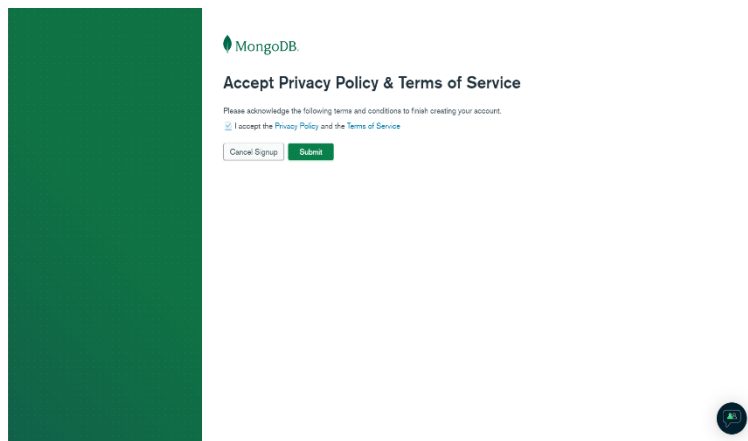
Accéder à la page de MongoDB cloud et cliquer sur bouton vert au milieu : La page est accessible au lien suivant [MongoDB Cloud | MongoDB](#)

Procurer les informations d'inscription pour un essai gratuit : Il est préférable de disposer d'un compte Google et de cliquer sur *se connecter avec Google*.

AIRLINE DASHBOARD

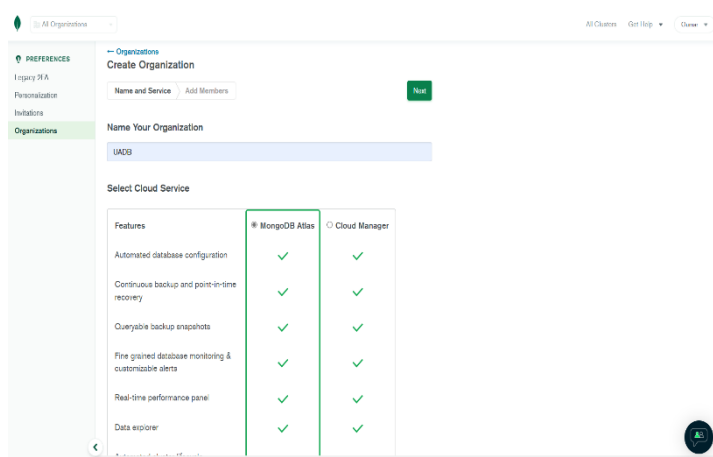
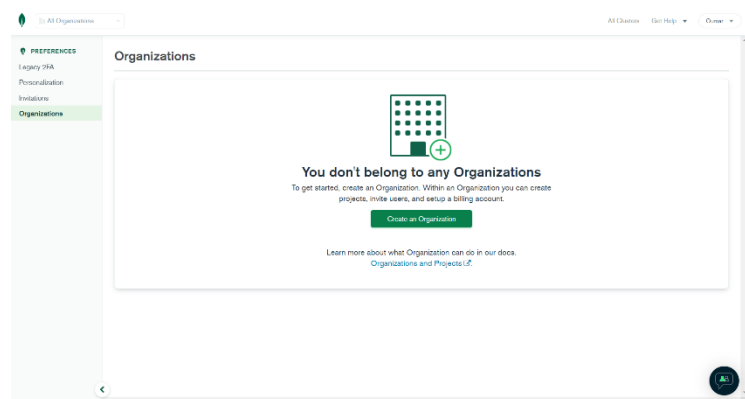
By  
Oumar Kane





Accepter les termes et conditions et valider pour terminer l'inscription : Cette étape succède à l'étape du choix du compte.

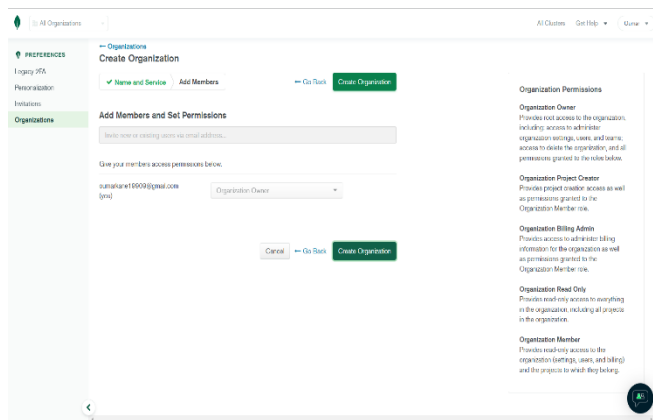
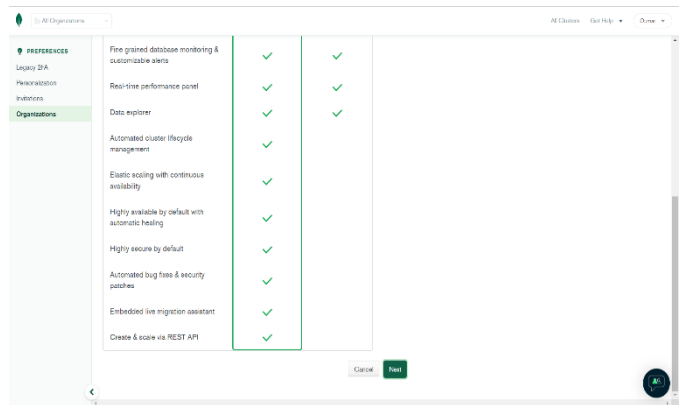
Cliquer sur le bouton *Créer une organisation* pour entamer la création d'une organisation.



Donner le nom de l'organisation : Les autres paramètres sont à laisser par défaut.

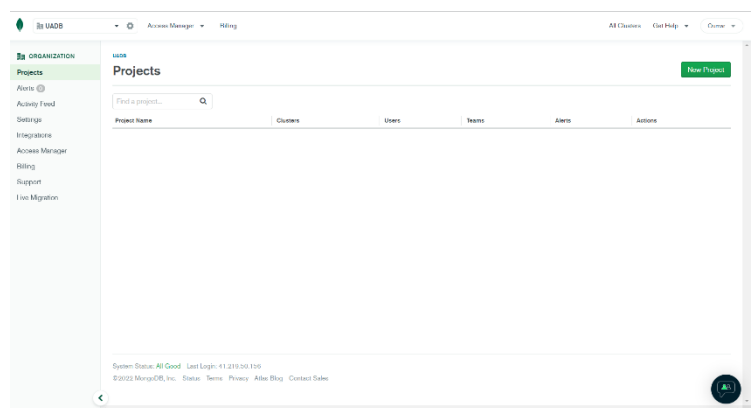
Cliquer sur le bouton *suivant* :

Cela nous permet d'accéder à la page suivante.



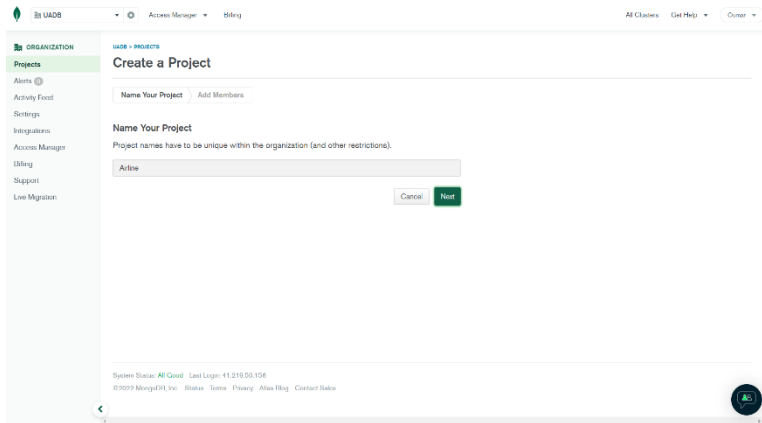
Cliquer sur le bouton *Créer l'organisation* pour valider la création : Il est possible d'ajouter des membres avant de valider.

Cliquer sur *créer un projet* à droite pour entamer la création d'un projet.



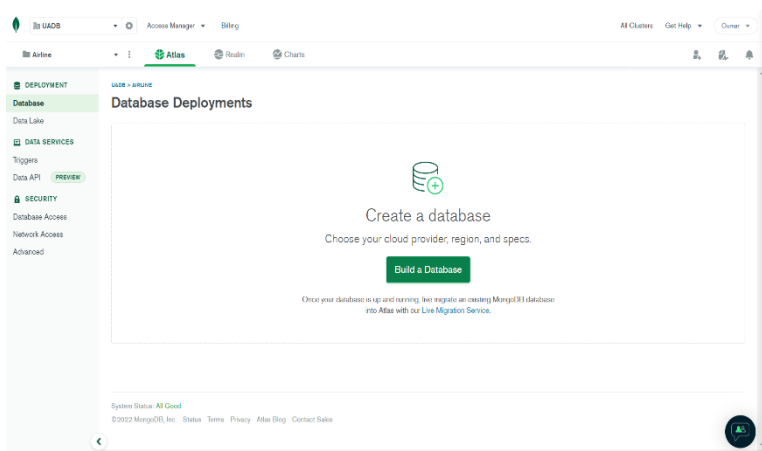
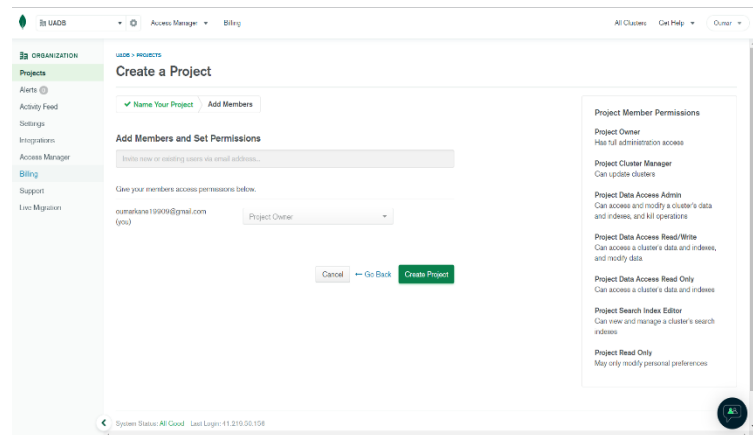
**AIRLINE DASHBOARD**

By  
Oumar Kane



Nommer le projet et cliquer sur *suivant* : Le nom du projet doit être unique (exemple : Airline).

Cliquer sur *créer le projet* pour valider la création du projet : Il est possible de permettre à des membres d'accéder au projet.

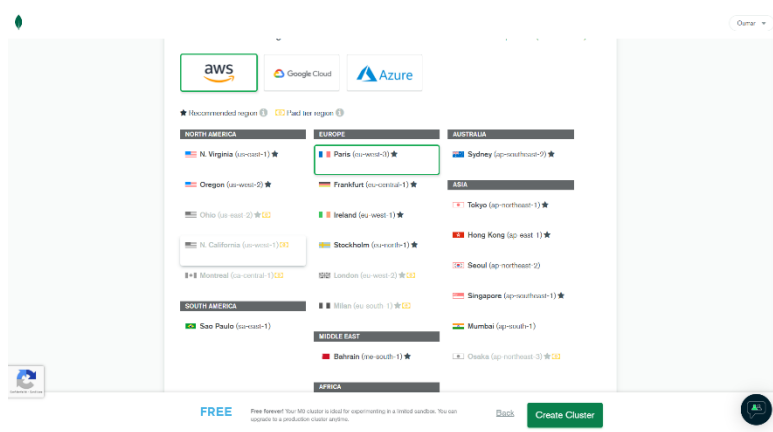
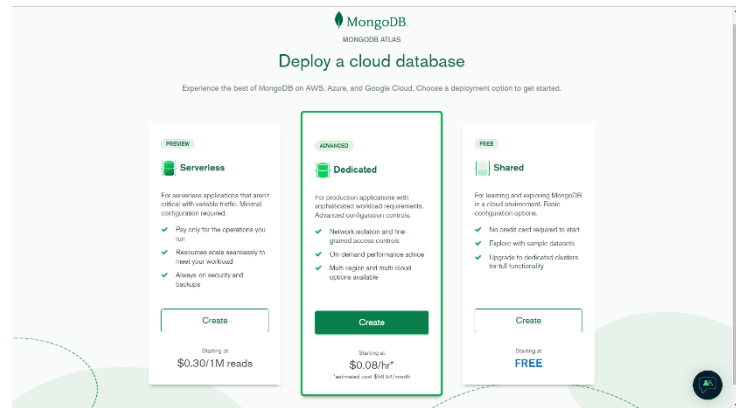


Cliquer sur *construire une base de données* pour entamer la conception d'une base de données (ou d'un cluster).

**AIRLINE DASHBOARD**

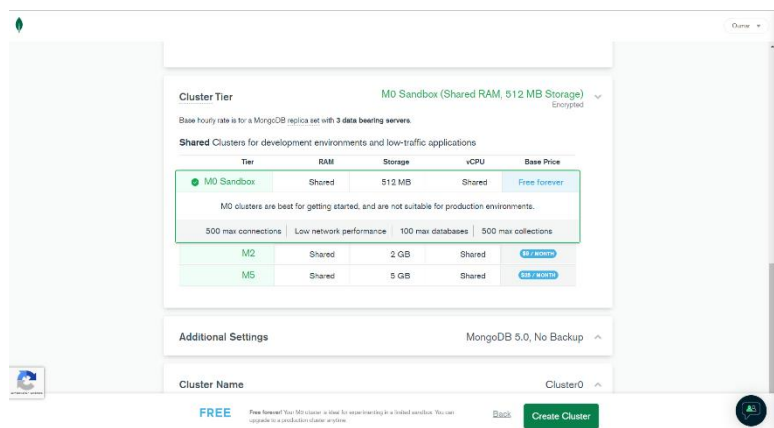
By  
Oumar Kane

Choisir le mode de déploiement *partagé* qui est gratuit et cliquer sur *créer* pour accéder à la page suivante.



Choisir le fournisseur de cloud et la région à partir de laquelle il sera fourni.

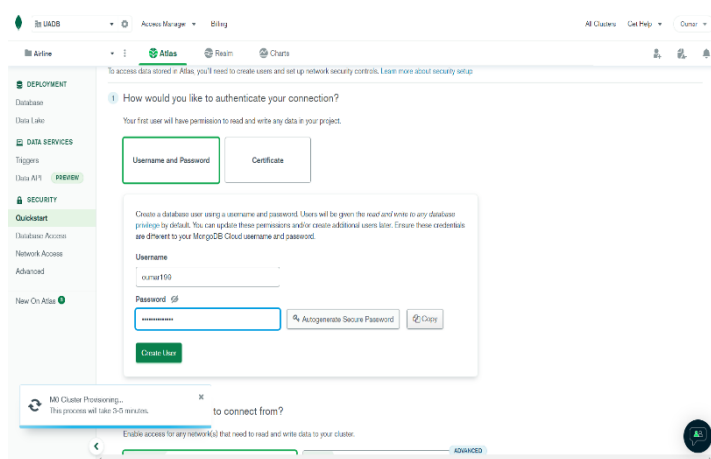
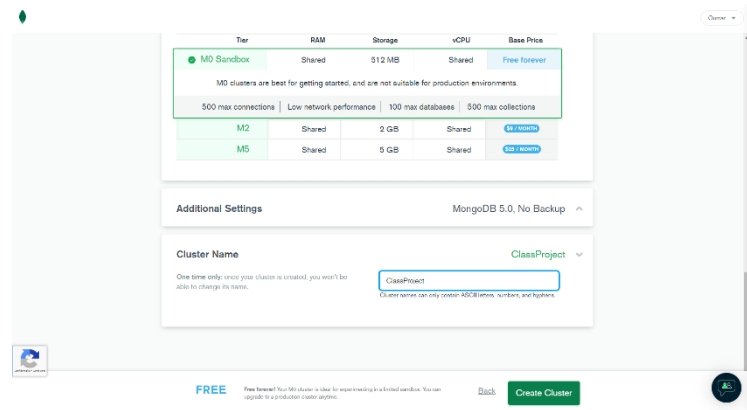
Choisir le type de stockage : Le type de stockage gratuit a pour taille maximale 512 MB.



**AIRLINE DASHBOARD**

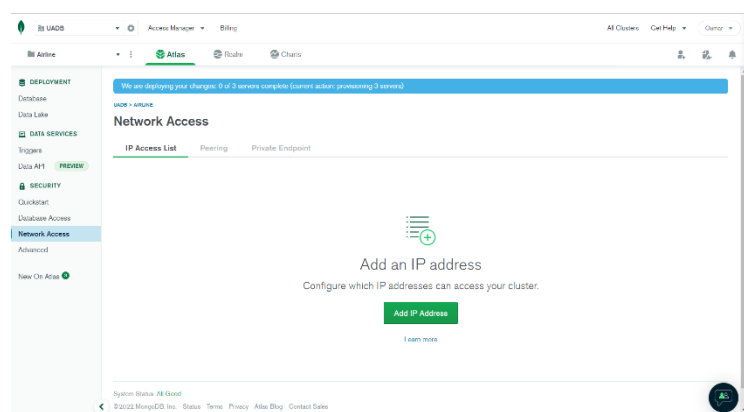
By  
Oumar Kane

Fournir le nom du cluster  
(Exemple : ClassProject)  
et cliquer sur *créer le cluster* pour valider la création.



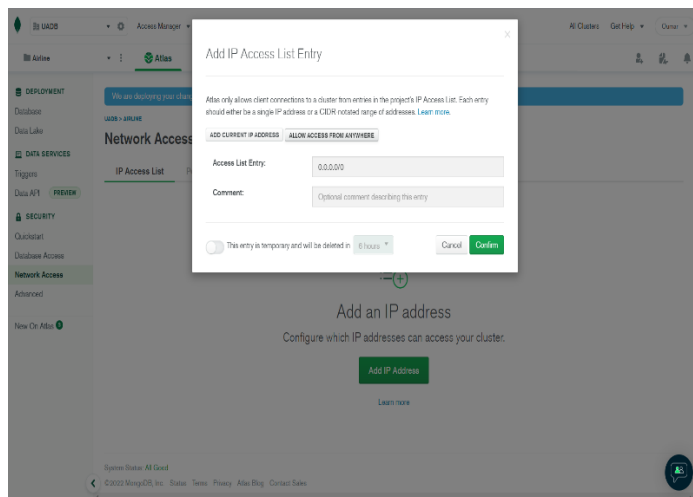
Fournir un nom d'utilisateur  
et un mot de passe et cliquer  
sur *créer l'utilisateur* : Le  
nom d'utilisateur ne  
contient pas d'espace  
(exemple : oumar199)

Sur la fenêtre Network  
Access ou accès réseau  
cliquer sur *ajouter une  
adresse IP (au milieu)*



AIRLINE DASHBOARD

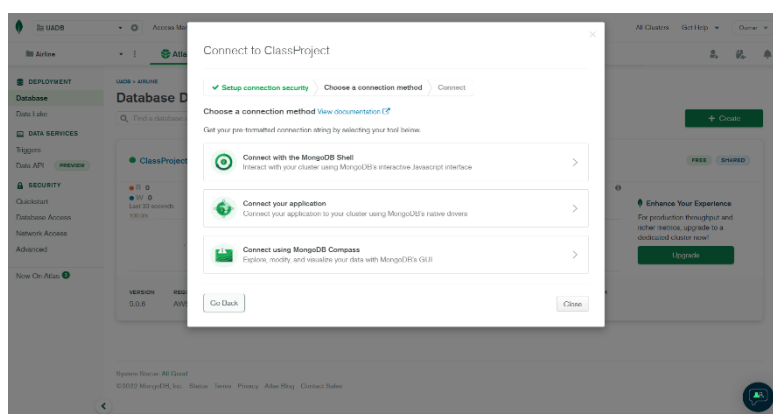
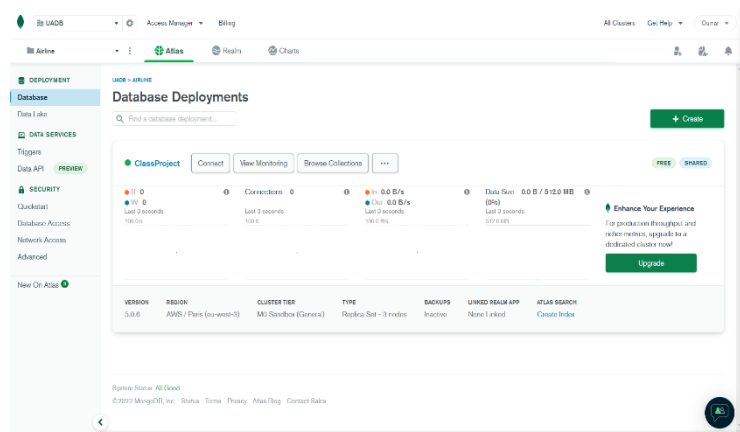
By  
Oumar Kane



Cliquer sur le deuxième bouton (*Accéder à partir de n'importe où*) et confirmer :

L'adresse IP mondiale n'est pas recommandée mais nous permet de nous connecter facilement au cluster.

Aller sur la fenêtre *Database* ou base de données et cliquer sur le bouton *Se connecter* à coté du nom du cluster.

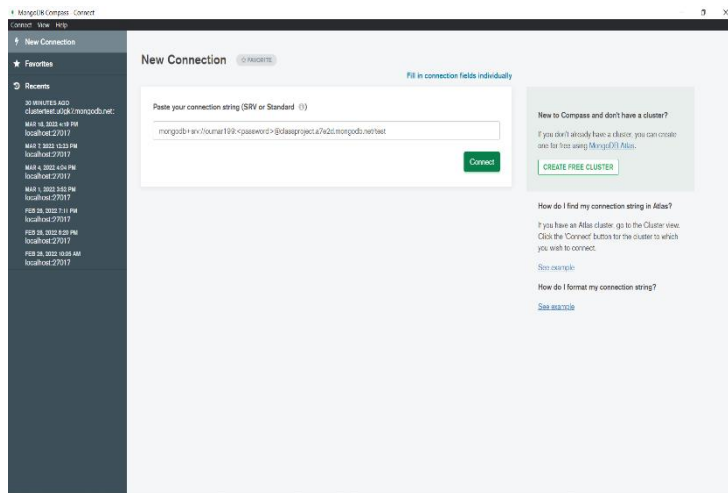
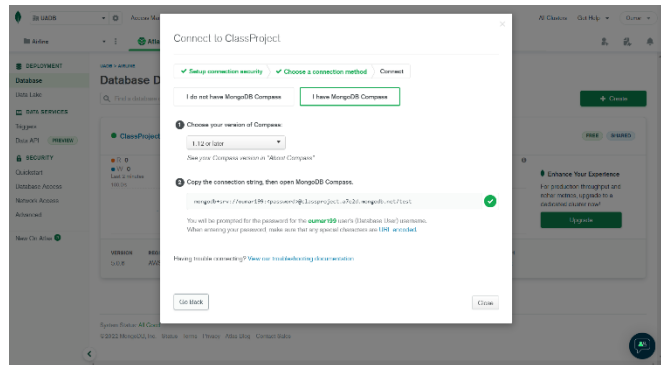


Cliquer sur *Se connecter avec MongoDB Compass* : On souhaite l'utiliser pour l'importation des données.

**AIRLINE DASHBOARD**

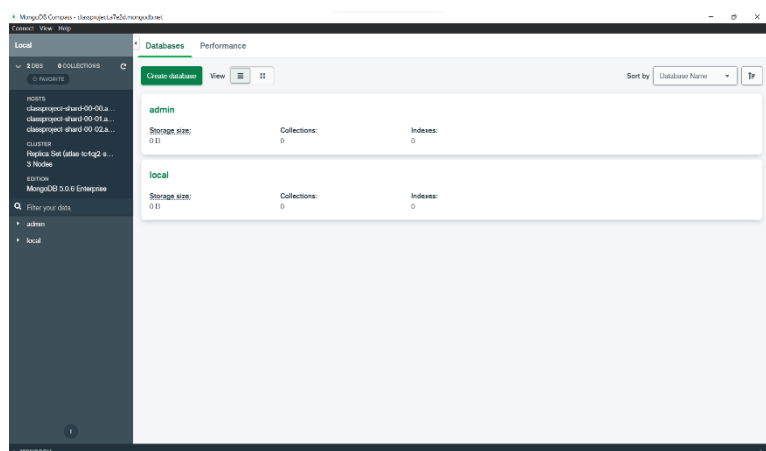
By  
Oumar Kane

Choisir *J'ai déjà MongoDB Compass*, copier le lien de connexion vers le serveur et cliquer sur *Fermer* pour fermer la fenêtre.



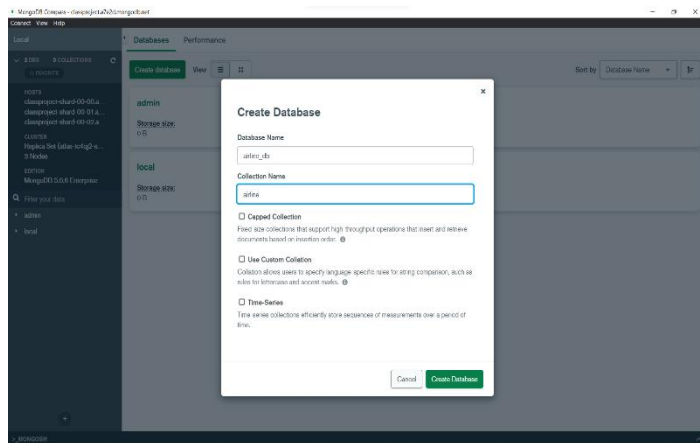
Ouvrir MongoDB Compass et coller le lien sur la barre de nouvelle connexion, remplacer <password> par son mot de passe et cliquer sur *Se connecter*.

Cliquer sur *Créer une base de données* : Pour procéder à la création d'une nouvelle base de données.



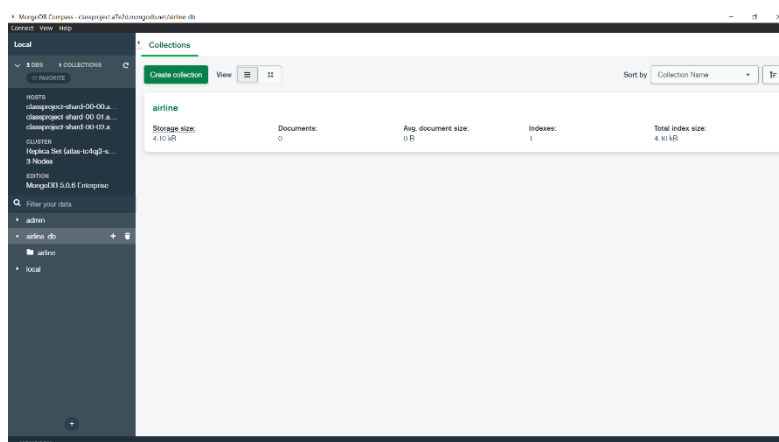
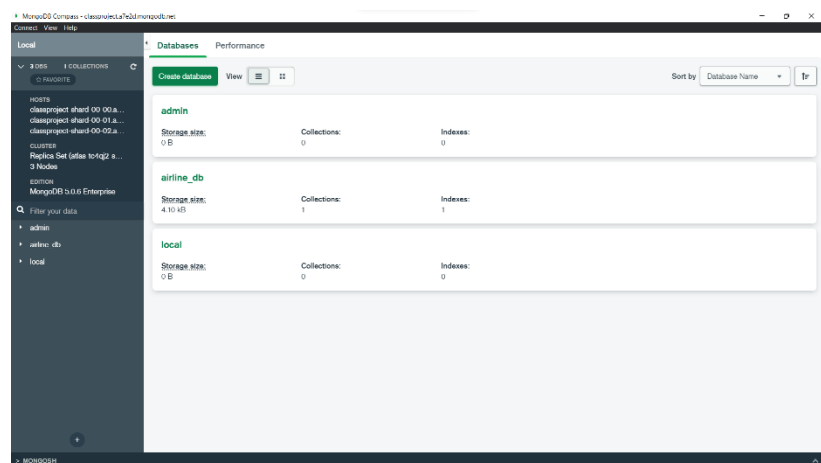
**AIRLINE DASHBOARD**

By  
Oumar Kane



Fournir le nom de la base de données ainsi que celui de la collection (exemple : `airline_db` et `airline` respectivement) et cliquer sur *créer la base de données*.

Sélectionner la base de données.



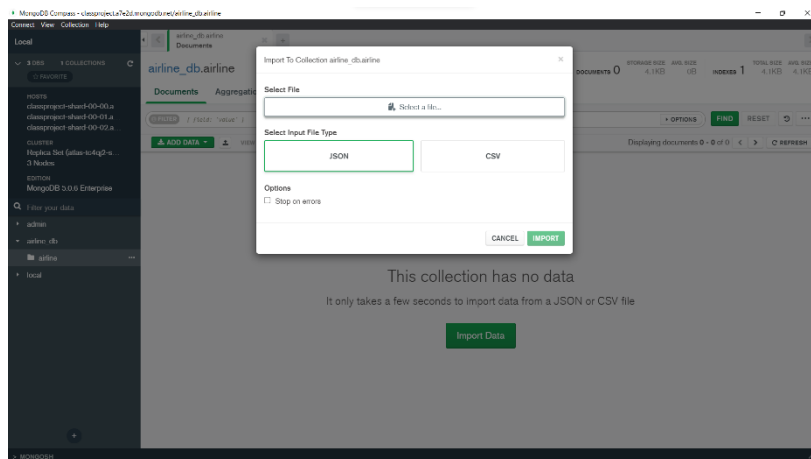
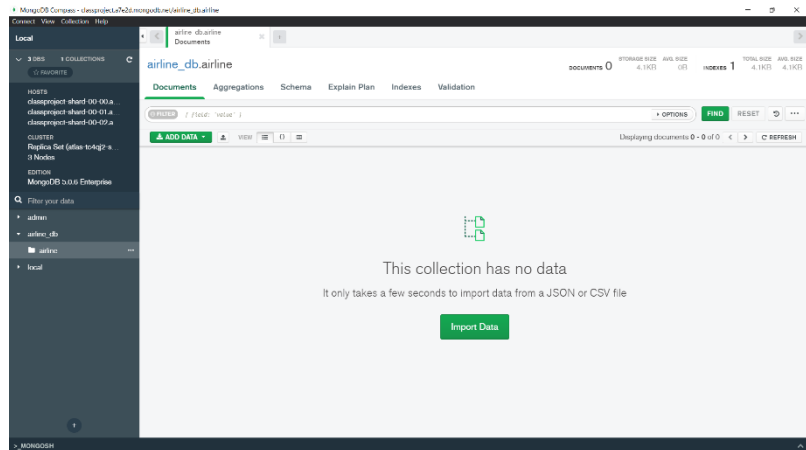
Sélectionner la collection.

**AIRLINE DASHBOARD**

By  
Oumar Kane

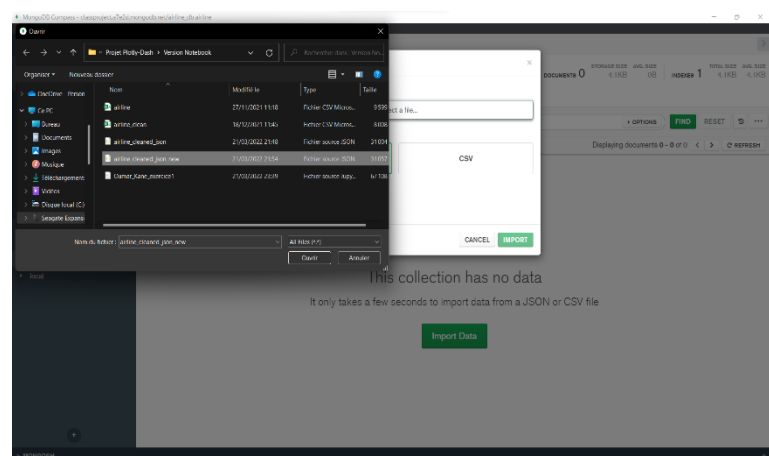


Cliquer sur *importer*  
*des données.*



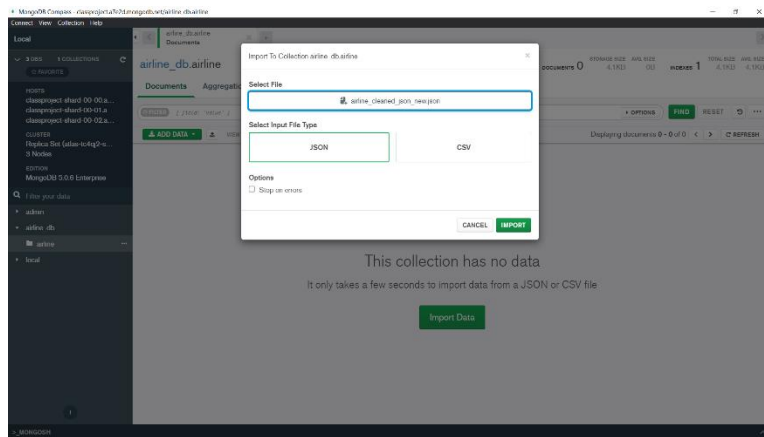
Sélectionner *JSON*  
pour le format du  
fichier et cliquer sur  
*Sélectionner un*  
*fichier.*

Sélectionner le fichier  
JSON à utiliser et cliquer  
sur *ouvrir.*



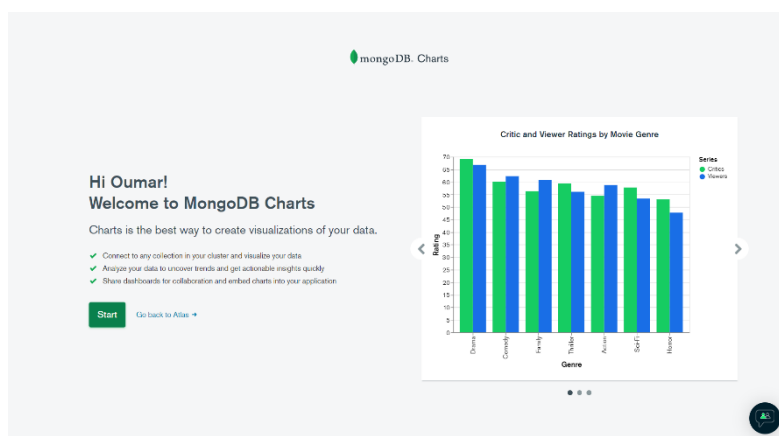
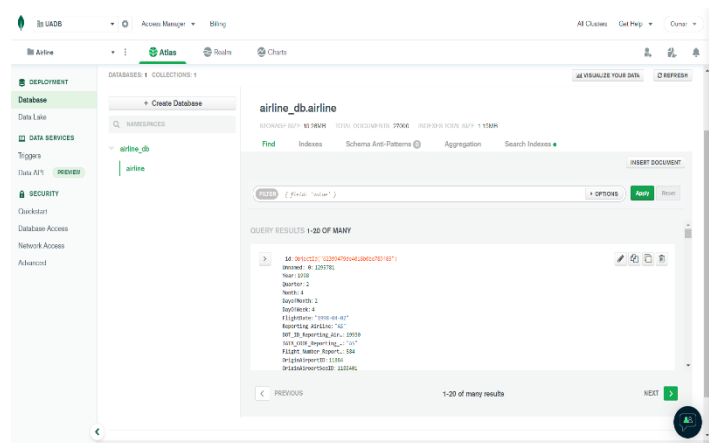
**AIRLINE DASHBOARD**

By  
Oumar Kane



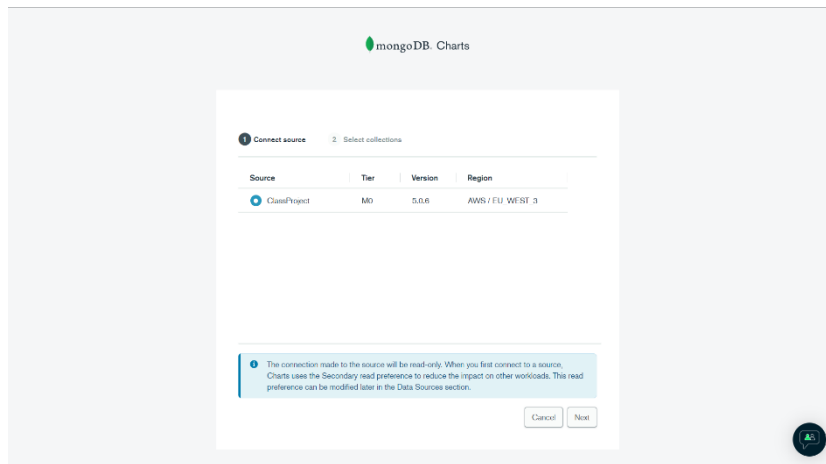
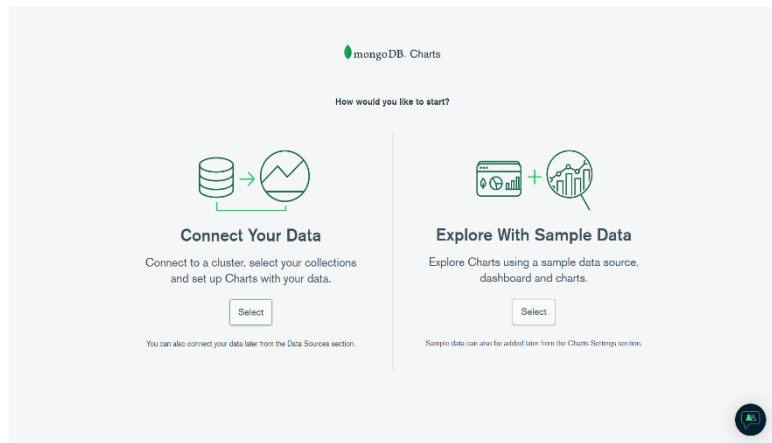
Cliquer sur *Importer* pour téléporter le fichier au niveau du cloud : Cela nécessite un certain temps suivant la taille du fichier.

Au niveau de MongoDB Atlas les données sont importées. Cliquer sur le bouton *Charts* (sur la barre horizontale) pour accéder à MongoDB Charts.



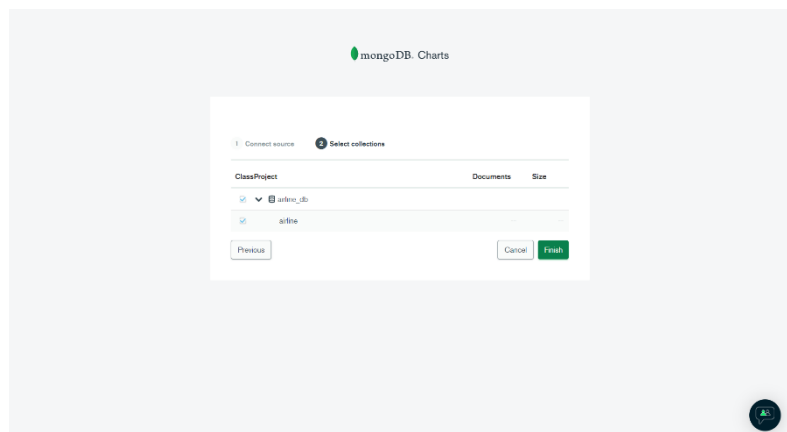
Sur la fenêtre de MongoDB Charts cliquer sur *Démarrer*.

Sélectionner *Connecter* vos données pour effectuer des tableaux de bord à partir du cluster.



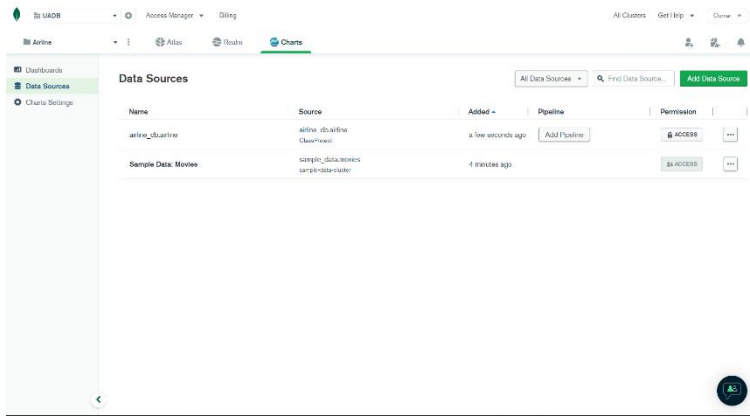
Sélectionner le cluster et cliquer sur le bouton *Suivant*.

Sélectionner la base de données et la collection et cliquer sur le bouton *Finir* pour terminer la procédure.



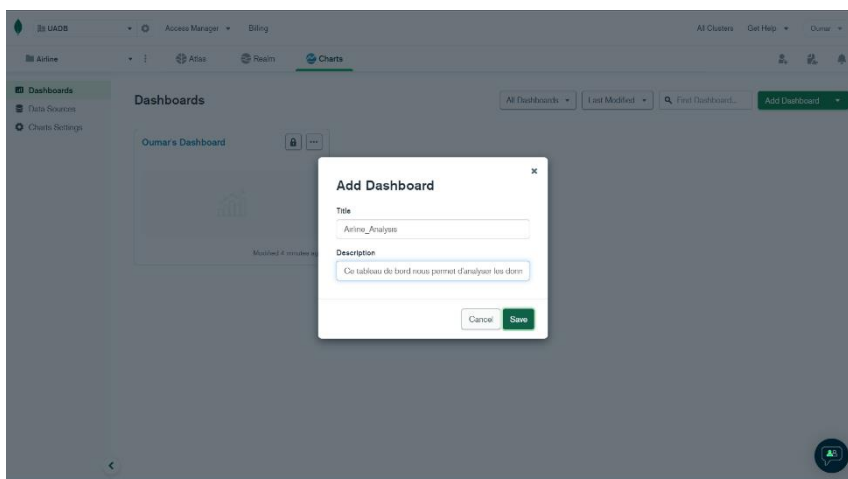
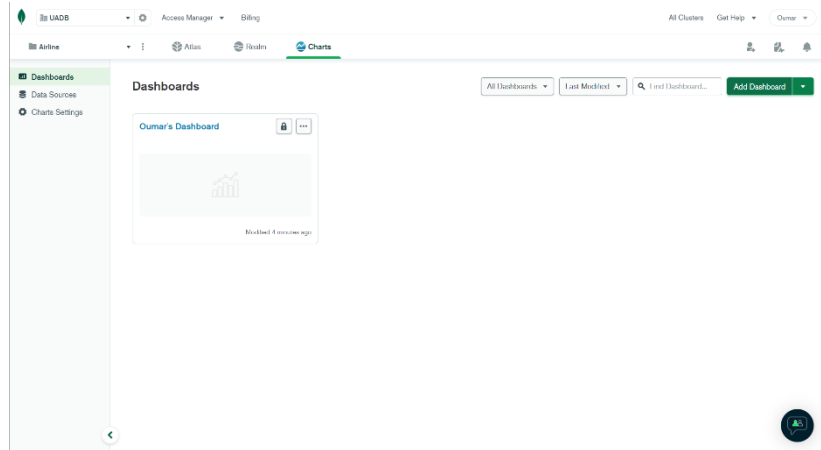
**AIRLINE DASHBOARD**

By  
Oumar Kane



Cliquer sur *Dashboards* ou tableaux de bord au niveau de l'onglet gauche.

Cliquer sur le bouton vert *Ajouter un tableau de bord.*

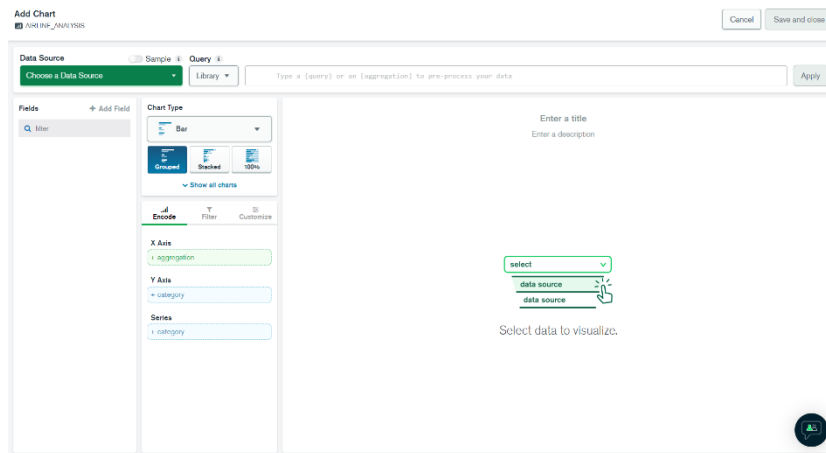
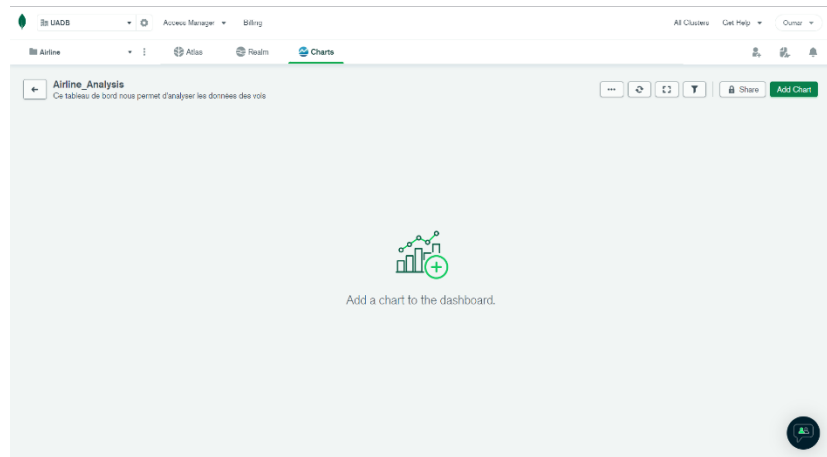


Ajouter le titre et la description du tableau de bord et cliquer sur *Sauvegarder.*

**AIRLINE DASHBOARD**

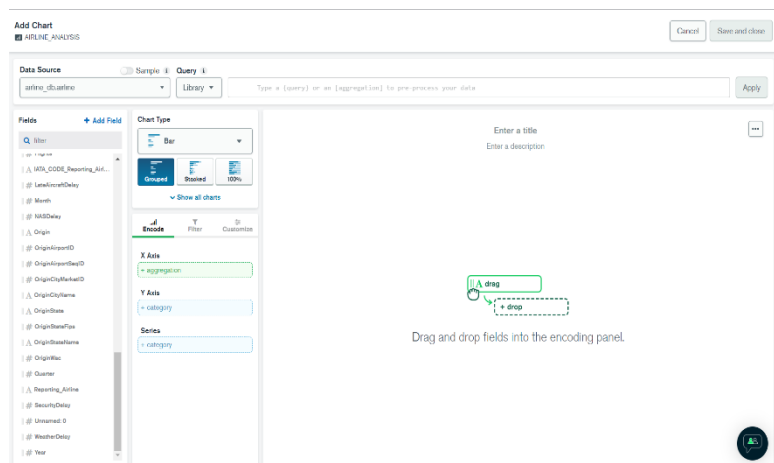
By  
Oumar Kane

Cliquer sur le bouton vert *Ajouter un graphique* : On peut en ajouter autant qu'on souhaite.



Cliquer sur *Choisir une source de données* pour sélectionner la collection à partir de laquelle sera créé le graphique.

Commencer à concevoir le graphique avec les outils disponibles.



**AIRLINE DASHBOARD**

By  
Oumar Kane

### III. DESCRIPTION DU JEU DE DONNEES

Le jeu de données utilisé est une version moins longue, nettoyée et mise sous format JSON du jeu de données Airlines Delay disponible sur [kaggle](#) à l'adresse suivante [airline\\_delay\\_causes](#).

Le jeu de données contient les informations sur 27 000 vols intérieurs effectués par de grands transporteurs aériens américains. Ces informations ont été fournies par le BTS (Bureau of Transportation Statistics) du département des Transports des Etats-Unis (DOT). Il renferme notamment des renseignements sommaires sur le nombre de vols à temps, retardés, annulés et déroutés figurant dans le Rapport sur les Consommateurs de Voyages Aériens effectué mensuellement par le DOT, généralement publié 30 jours après la fin d'un mois sous forme de tableaux sommaires.

Les données recueillies vont nous permettre d'effectuer une analyse sur les retards de plusieurs vols répertoriés sur le territoire américain, de 1987 à 2020 et nous permettre ainsi d'y apporter des explications.

Plusieurs variables vont entrer en jeu dans nos analyses. Il est ainsi crucial, avant de débiter, de décrire brièvement les principaux attributs de notre jeu de données :

- **Year (entier)** : L'Année durant laquelle le vol s'est déroulé. Les valeurs sont comprises entre 1987 et 2020 ;
- **Quarter (entier)** : Le trimestre durant lequel le vol s'est déroulé. Les valeurs sont comprises entre 1 et 4 ;
- **Month (entier)** : Le mois durant lequel le vol s'est produit. Les valeurs sont comprises entre 1 (Janvier) et 12 (Décembre) ;
- **DayofMonth (entier)** : Le jour du mois durant lequel le vol s'est déroulé. Les valeurs sont comprises entre 1 et 31 ;
- **DayOfWeek (entier)** : Le jour de la semaine durant lequel le vol s'est déroulé. Les valeurs sont comprises entre 1 (lundi) et 7 (dimanche) ;
- **FlightDate (date)** : La date du vol en format année-mois-jour ;
- **Reporting\_Airline (chaîne de caractères)** : Le code de la compagnie aérienne en charge du vol ;
- **OriginAirportID (entier)** : L'identifiant de l'aéroport de départ ;
- **Origin (chaîne de caractères)** : Le code de la ville de départ ;
- **OriginCityName (chaîne de caractères)** : Le nom de la ville de départ ;
- **OriginState (chaîne de caractères)** : Le code de l'Etat d'origine ;
- **OriginStateName (chaîne de caractères)** : Le nom de l'Etat d'origine ;
- **DestAirportID (entier)** : L'identifiant de l'aéroport de destination ;

- **Dest (chaîne de caractères)** : Le code de la ville de destination ;
- **DestCityName (chaîne de caractères)** : Le nom de la ville de destination ;
- **DestState (chaîne de caractères)** : Le code de l'Etat de destination ;
- **DestStateName (chaîne de caractères)** : Le nom de l'Etat de destination ;
- **DepDelayMinutes (décimale)** : Le retard au départ du vol en nombre de minutes ;
- **ArrDelayMinutes (décimale)** : Le retard à l'arrivée du vol en nombre de minutes ;
- **Cancelled (entier)** : Indique si le vol a été annulé (1) ou non (0) ;
- **Diverted (entier)** : Indique si le vol a été détourné (1) ou non (0) ;
- **ActualElapsedTime (décimale)** : Le temps de vol total ;
- **Flights (entier)** : Permet de compter le vol (ne contient que des 1) ;
- **Distance (décimale)** : La distance parcourue lors du vol ;
- **CarrierDelay (décimale)** : Le retard du transporteur en nombre de minutes ;
- **WeatherDelay (décimale)** : Le retard météorologique en nombre de minutes ;
- **NASDelay (décimale)** : Le retard du système aérien national en nombre de minutes ;
- **SecurityDelay (décimale)** : Le retard de la sécurité en nombre de minutes ;
- **LateAircraftDelay (décimale)** : Le retard des avions en nombre de minutes.

## IV. MISE EN ŒUVRE DES GRAPHIQUES

Les graphiques à concevoir doivent nous permettre d'analyser les données sur les retards des vols aux Etats-Unis des points de vue géographique et temporel. Ainsi, nous allons répartir les graphiques selon deux groupes différents : Ceux essentiellement basés sur les Etats et les villes des Etats-Unis, puis ceux permettant de déterminer les compagnies aériennes qui enregistrent les plus grands nombres de retards au fil des années (cet analyse portera plus sur les années entre 2003 et 2020).

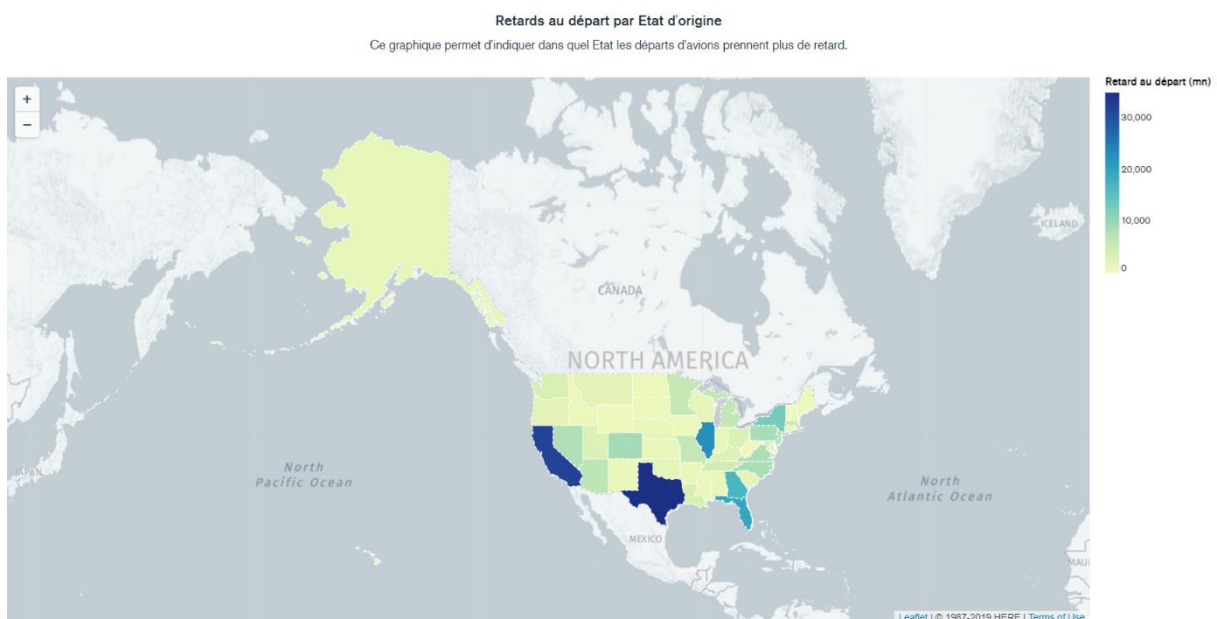
### 1. Du point de vue Géographique

#### a. Retards par Etat

Les premiers graphiques permettent de déterminer les Etats qui ont enregistré le plus grand nombre de retards à l'arrivée ou au départ d'avions.

##### - Retards au départ par Etat d'origine

Ce graphique permet d'indiquer dans quel Etat les départs d'avions prennent plus de retard.





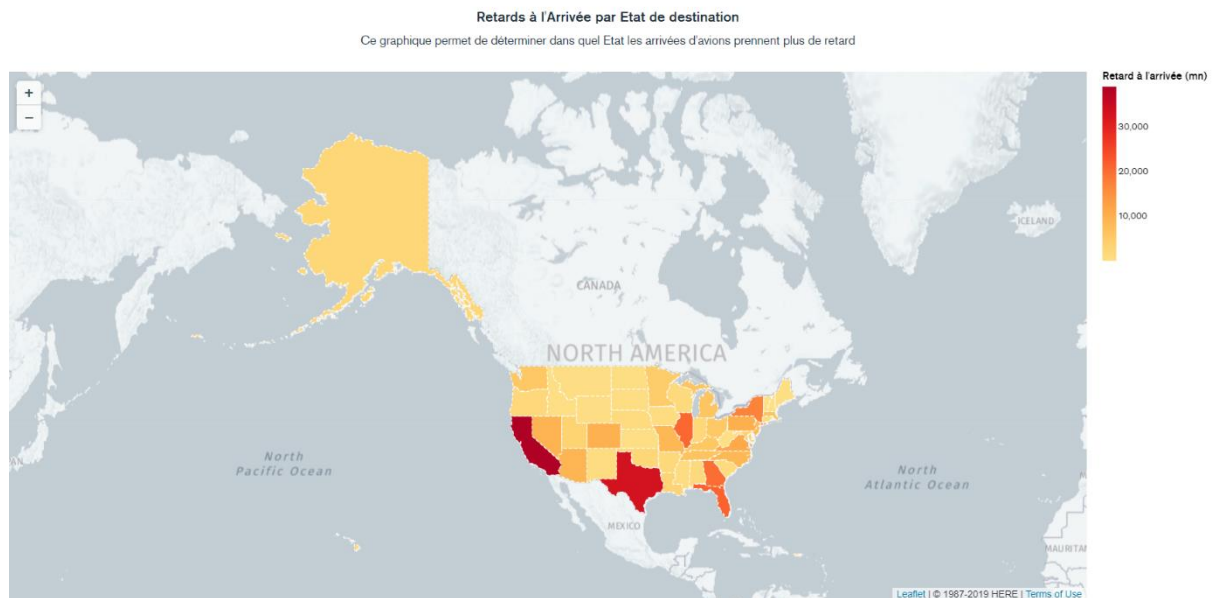
**Analyse :** Les Etats enregistrant le plus de retards au départ d'avions sont principalement le Texas (en première position) et la Californie (en seconde position).

**Pipeline d'agrégation :**

```
[
  {
    "$group": {
      "_id": {
        "__alias_0": "$OriginStateName"
      },
      "__alias_1": {
        "$sum": "$DepDelayMinutes"
      }
    }
  },
  {
    "$project": {
      "_id": 0,
      "__alias_0": "$_id.__alias_0",
      "__alias_1": 1
    }
  },
  {
    "$project": {
      "location": "$__alias_0",
      "color": "$__alias_1",
      "_id": 0
    }
  },
]
```

## - Retards à l'arrivée par Etat de destination

Ce graphique permet de déterminer dans quel Etat les arrivées d'avions prennent plus de retard.



**Analyse :** Les Etats enregistrant le plus de retards à l'arrivée sont principalement la Californie (en première position) et le Texas (en seconde position).

## Pipeline d'agrégation :

```
[
  {
    "$group": {
      "_id": {
        "__alias_0": "$DestStateName"
      },
      "__alias_1": {
        "$sum": "$ArrDelayMinutes"
      }
    }
  },
  {
    "$project": {
      "_id": 0,
      "__alias_0": "$_id.__alias_0",
      "__alias_1": 1
    }
  },
  {
    "$project": {
      "color": "$__alias_1",
      "location": "$__alias_0",
      "_id": 0
    }
  },
  {
    "$limit": 5000
  }
]
```

**Conclusion :** Aussi bien pour les départs que pour les arrivées, le Texas et la Californie enregistrent le plus grand nombre de retards. Cela peut être dû au fait que ces Etats sont les moins accessibles aux Etats-Unis à cause de la longue distance à parcourir et que le plus grand nombre de vols est effectué au niveau de

ces deux Etats. Pour visualiser l'analyse de la distance et du nombre de vols par Etat d'origine et de destination, voire le tableau de bord final obtenu.

### b. Nombres d'annulations et de déviations par Etat

Les graphiques des nombres d'annulations et ceux des nombres de déviations par Etat d'origine et de destination nous permettront de connaître les Etats qui enregistrent les plus grands nombres d'annulations et/ou de déviations de vols.

#### - Calcul des nombres totaux d'annulations et de déviations

Nombre d'annulations  
Calcul du nombre total d'annulations

# 447

Nombre total de déviations  
Calcul du nombre total de déviations

# 55

## Pipelines d'agrégation respectifs

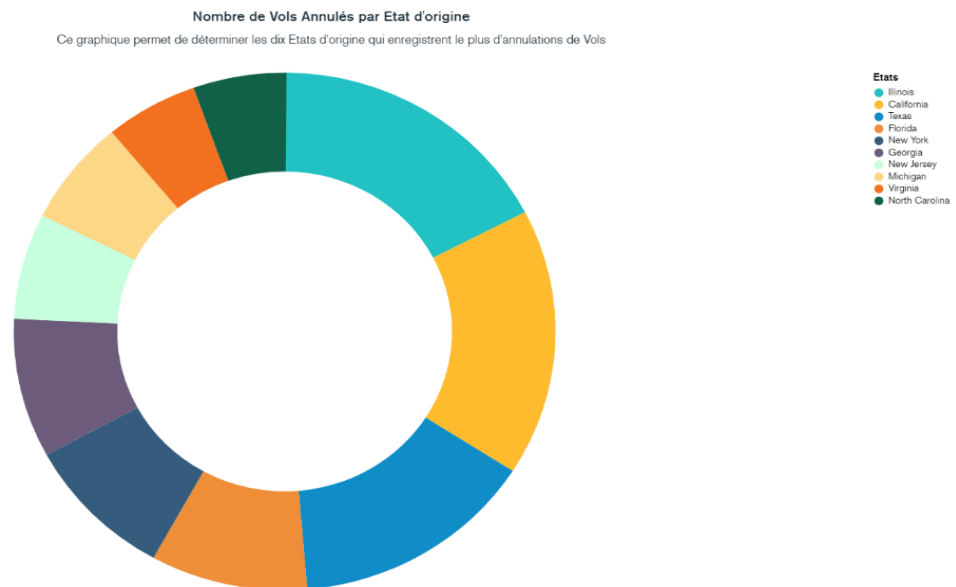
```
[
  {
    "$group": {
      "_id": {},
      "__alias_0": {
        "$sum": "$Cancelled"
      }
    }
  },
  {
    "$project": {
      "_id": 0,
      "__alias_0": 1
    }
  },
  {
    "$project": {
      "value": "$__alias_0",
      "_id": 0
    }
  },
  {
    "$limit": 5000
  }
]
```

```
[
  {
    "$group": {
      "_id": {},
      "__alias_0": {
        "$sum": "$Diverted"
      }
    }
  },
  {
    "$project": {
      "_id": 0,
      "__alias_0": 1
    }
  },
  {
    "$project": {
      "value": "$__alias_0",
      "_id": 0
    }
  },
  {
    "$limit": 5000
  }
]
```

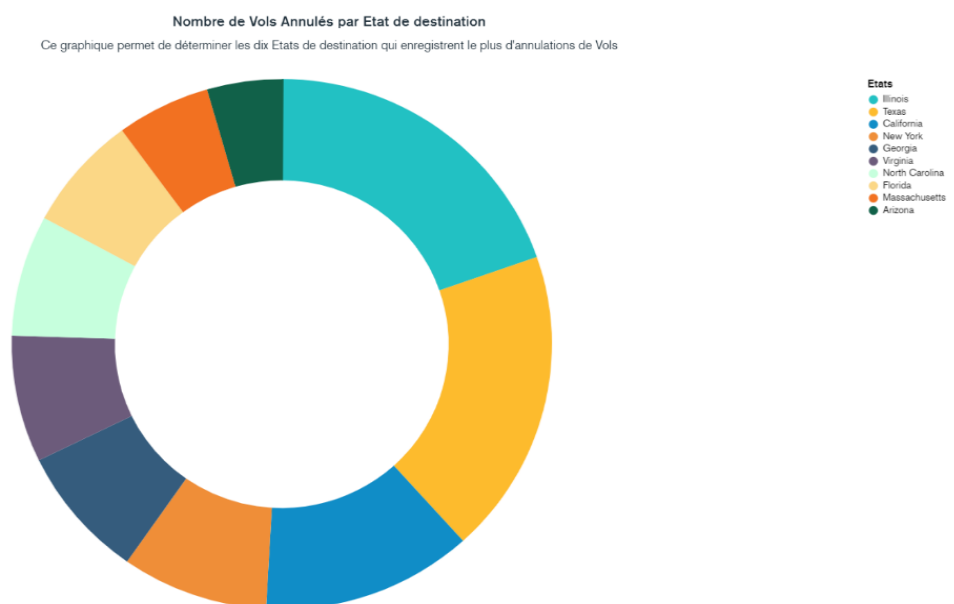
## - Nombres de Vols annulés par Etat d'origine et par Etat de destination

Les graphiques suivants permettent de déterminer les Etats qui enregistrent les plus grands nombres d'annulations de vol.

### ➤ Nombre d'annulations par Etat d'origine



### ➤ Nombre d'annulations par Etat de destination



**AIRLINE DASHBOARD**

By  
Oumar Kane

**Analyse :** Nous remarquons, sur les deux graphiques, que les Etats qui enregistrent les plus grands nombres d’annulations sont principalement l’Illinois, Le Texas et la Californie.

**Pipeline d’agrégation du nombre d’annulations par Etat d’origine et celui du nombre d’annulations par Etat de destination :**

```
[
  {
    "$group": {
      "_id": {
        "__alias_0": "$OriginStateName"
      },
      "__alias_1": {
        "$sum": "$Cancelled"
      }
    }
  },
  {
    "$project": {
      "_id": 0,
      "__alias_0": "$_id.__alias_0",
      "__alias_1": 1
    }
  },
  {
```

```
    "$project": {
      "value": "$__alias_1",
      "label": "$__alias_0",
      "_id": 0
    }
  },
  {
    "$addFields": {
      "__agg_sum": {
        "$sum": [
          "$value"
        ]
      }
    }
  },
  {
    "$sort": {
      "__agg_sum": -1
    }
  }
```

```
},
{
  "$limit": 10
},
{
  "$project": {
    "__agg_sum": 0
  }
},
{
  "$limit": 5000
}
]
```

```
[
  {
    "$group": {
      "_id": {
        "__alias_0": "$DestStateName"
      },
      "__alias_1": {
        "$sum": "$Cancelled"
      }
    }
  },
  {
    "$project": {
      "_id": 0,
      "__alias_0": "$_id.__alias_0",
      "__alias_1": 1
    }
  },
  {
```

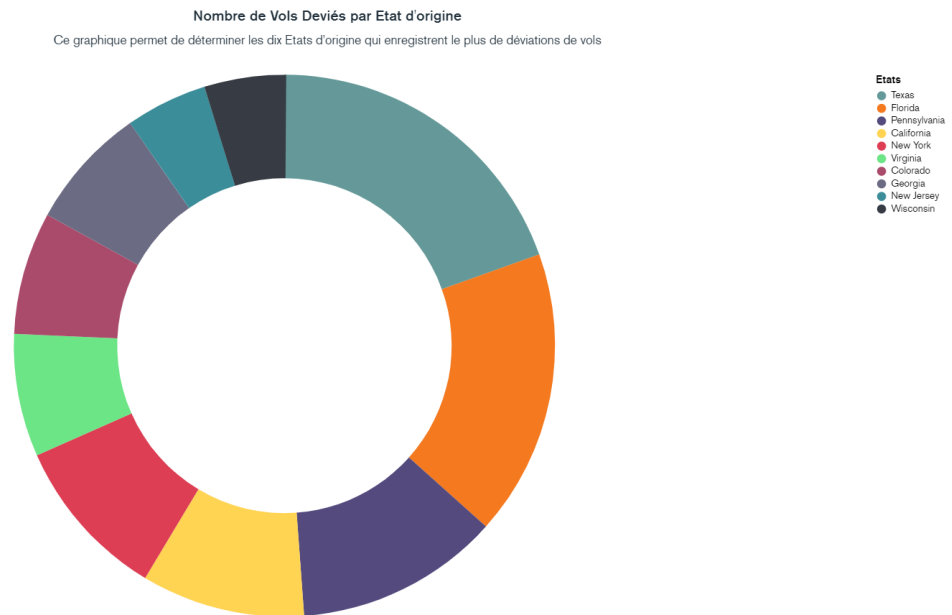
```
    "$project": {
      "value": "$__alias_1",
      "label": "$__alias_0",
      "_id": 0
    }
  },
  {
    "$addFields": {
      "__agg_sum": {
        "$sum": [
          "$value"
        ]
      }
    }
  },
  {
    "$sort": {
      "__agg_sum": -1
    }
  }
```

```
},
{
  "$limit": 10
},
{
  "$project": {
    "__agg_sum": 0
  }
},
{
  "$limit": 5000
}
]
```

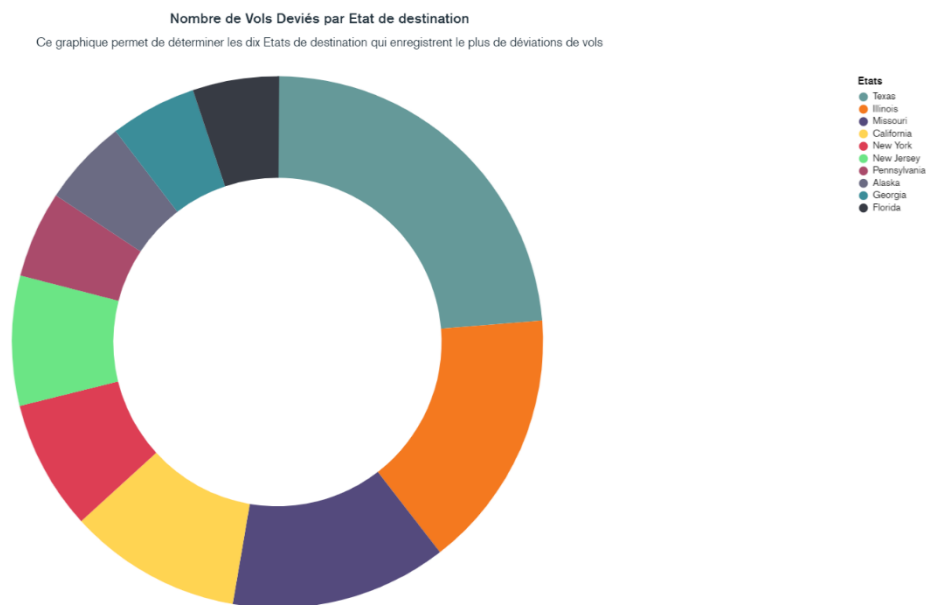
## - Nombres de Vols déviés par Etat d'origine et par Etat de destination

Les graphiques suivants permettent de déterminer les Etats qui enregistrent les plus grands nombres de déviations de vol.

### ➤ Nombre de déviations par Etat d'origine



### ➤ Nombre de déviations par Etat de destination



**AIRLINE DASHBOARD**

By  
Oumar Kane



**Analyse :** Nous remarquons, sur les deux graphiques, que le Texas est l'Etat qui enregistre le plus grand nombre de déviations d'avions sur le territoire américain.

Pipeline d'agrégation du nombre de déviations par Etat d'origine et celui du nombre de déviations par Etat de destination :

```
[
  {
    "$group": {
      "_id": {
        "__alias_0": "$OriginStateName"
      },
      "__alias_1": {
        "$sum": "$Diverted"
      }
    }
  },
  {
    {
      "$project": {
        "_id": 0,
        "__alias_0": "$_id.__alias_0",
        "__alias_1": 1
      }
    }
  },
  {
```

```
    "$project": {
      "value": "$__alias_1",
      "label": "$__alias_0",
      "_id": 0
    }
  },
  {
    {
      "$addFields": {
        "__agg_sum": {
          "$sum": [
            "$value"
          ]
        }
      }
    }
  },
  {
    "$sort": {
      "__agg_sum": -1
    }
  }
```

```
},
{
  "$limit": 10
},
{
  {
    "$project": {
      "__agg_sum": 0
    }
  },
  {
    "$limit": 5000
  }
}
]
```

```
[
  {
    "$group": {
      "_id": {
        "__alias_0": "$DestStateName"
      },
      "__alias_1": {
        "$sum": "$Diverted"
      }
    }
  },
  {
    {
      "$project": {
        "_id": 0,
        "__alias_0": "$_id.__alias_0",
        "__alias_1": 1
      }
    }
  },
  {
```

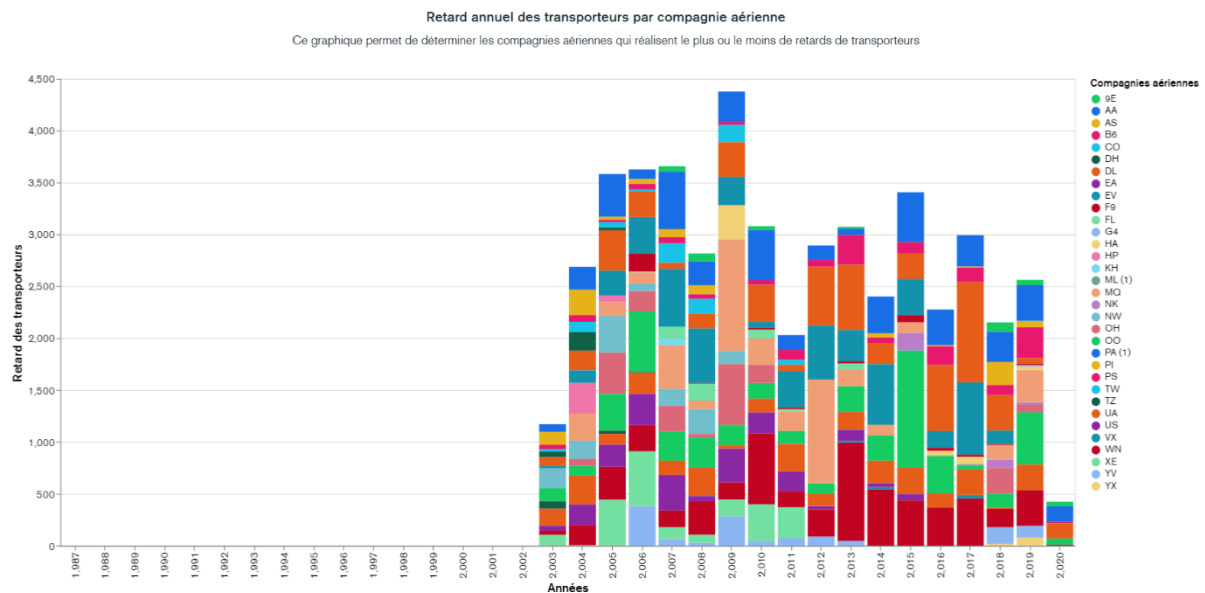
```
    "$project": {
      "value": "$__alias_1",
      "label": "$__alias_0",
      "_id": 0
    }
  },
  {
    {
      "$addFields": {
        "__agg_sum": {
          "$sum": [
            "$value"
          ]
        }
      }
    }
  },
  {
    "$sort": {
      "__agg_sum": -1
    }
  }
```

```
},
{
  "$limit": 10
},
{
  {
    "$project": {
      "__agg_sum": 0
    }
  },
  {
    "$limit": 5000
  }
}
]
```

## 2. Du point de vue Temporel

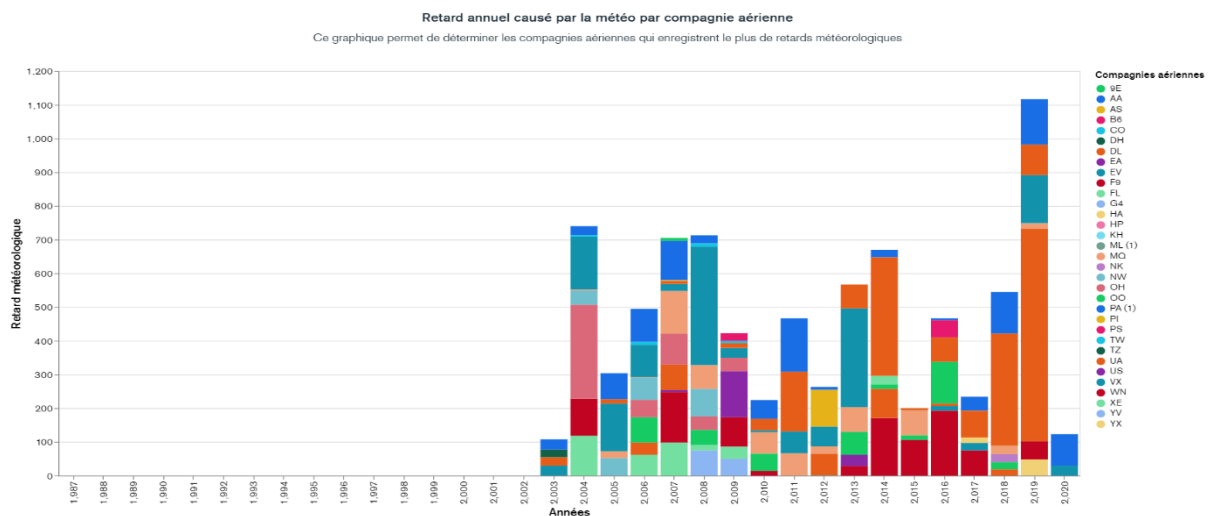
Il est temps de passer à l'analyse des retards annuels par compagnie aérienne. Ça nous permettra de déterminer les compagnies aériennes qui enregistrent le plus de retards de transporteurs, météorologiques, du système aérien national et/ou d'avions.

### ➤ Retard annuel de transporteurs par Compagnie Aérienne déclarante



Les 3 Compagnies aériennes avec le plus de retards de transporteurs sur les dernières années : AA, B6 et DL.

### ➤ Retard annuel météorologique par Compagnie Aérienne déclarante

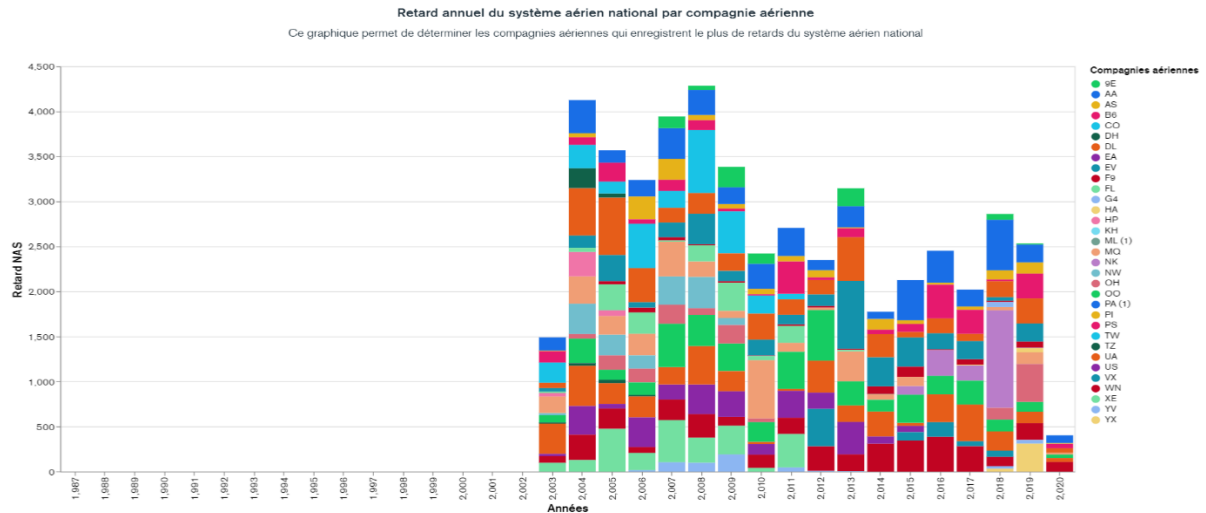


AIRLINE DASHBOARD

By  
Oumar Kane

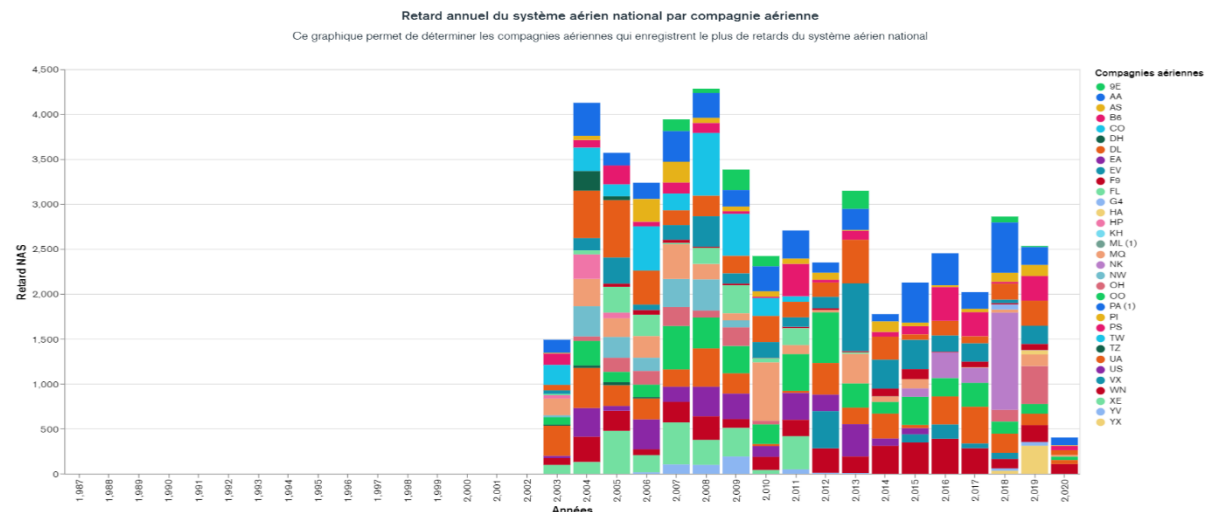
## Les 3 Compagnies aériennes avec le plus de retards météorologiques sur les dernières années : AA, DL et EV.

### ➤ Retard annuel du système aérien national par Compagnie Aérienne déclarante



## Les 3 Compagnies aériennes avec le plus de retards du NAS sur les dernières années : 9 E, AA et AS.

### ➤ Retard annuel d'avion par Compagnie Aérienne déclarante



## Les 3 Compagnies aériennes avec le plus de retards d'avion sur les dernières années : 9 E, AA et AS.

**Analyse :** En résumé nous pouvons dire que le top 3 des compagnies aériennes les moins sûres est le suivant : AA (en première position), 9 E (en seconde position) et DL (en troisième position).

**Pipeline d'agrégation du retard annuel de transporteurs par compagnie aérienne et celui du retard météorologique par compagnie aérienne**

```
[
  {
    "$group": {
      "_id": {
        "__alias_0": "$Year",
        "__alias_1": "$Reporting_Airline"
      },
      "__alias_2": {
        "$sum": "$CarrierDelay"
      }
    }
  },
  {
    "$project": {
      "_id": 0,
      "__alias_0": "$_id.__alias_0",
      "__alias_1": "$_id.__alias_1",
      "__alias_2": 1
    }
  }
],
```

```
{
  "$project": {
    "x": "$__alias_0",
    "y": "$__alias_2",
    "color": "$__alias_1",
    "_id": 0
  }
},
{
  "$group": {
    "_id": {
      "x": "$x"
    },
    "__grouped_docs": {
      "$push": "$$ROOT"
    }
  }
},
{
  "$sort": {
    "_id.x": 1
  }
},
{
  "$unwind": "$__grouped_docs"
},
{
  "$replaceRoot": {
    "newRoot": "$__grouped_docs"
  }
},
{
  "$limit": 5000
}
]
```

```
[
  {
    "$group": {
      "_id": {
        "__alias_0": "$Year",
        "__alias_1": "$Reporting_Airline"
      },
      "__alias_2": {
        "$sum": "$WeatherDelay"
      }
    }
  },
  {
    "$project": {
      "_id": 0,
      "__alias_0": "$_id.__alias_0",
      "__alias_1": "$_id.__alias_1",
      "__alias_2": 1
    }
  }
],
```

```
{
  "$project": {
    "x": "$__alias_0",
    "y": "$__alias_2",
    "color": "$__alias_1",
    "_id": 0
  }
},
{
  "$group": {
    "_id": {
      "x": "$x"
    },
    "__grouped_docs": {
      "$push": "$$ROOT"
    }
  }
},
{
  "$sort": {
    "_id.x": 1
  }
},
{
  "$unwind": "$__grouped_docs"
},
{
  "$replaceRoot": {
    "newRoot": "$__grouped_docs"
  }
},
{
  "$limit": 5000
}
]
```

**AIRLINE DASHBOARD**

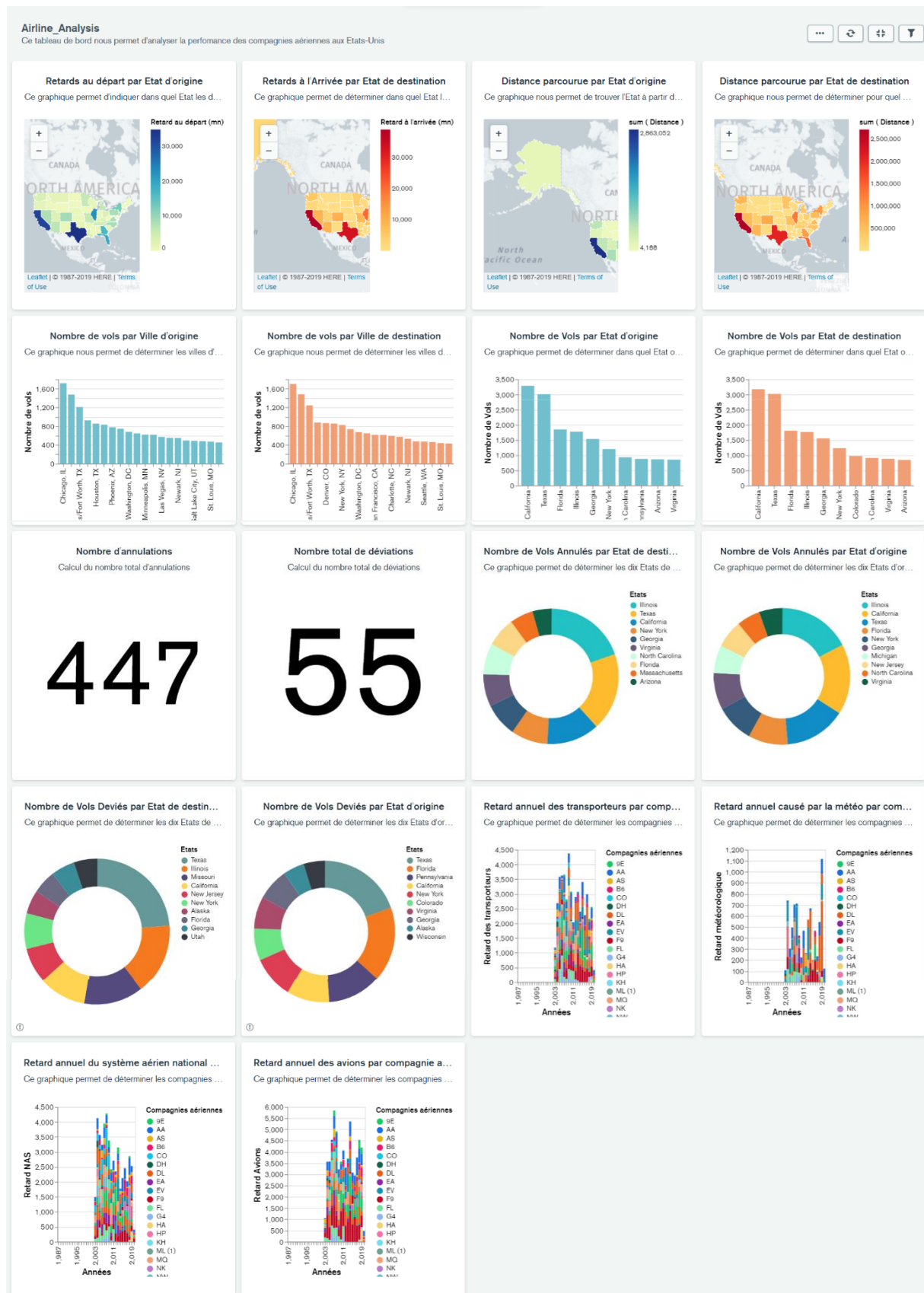
By  
Oumar Kane

## Pipeline d'agrégation du retard annuel du NAS par compagnie aérienne et celui du retard d'avions par compagnie aérienne

```
[
  {
    "$group": {
      "_id": {
        "__alias_0": "$Year",
        "__alias_1": "$Reporting_Airline"
      },
      "__alias_2": {
        "$sum": "$NASDelay"
      }
    }
  },
  {
    {
      "$project": {
        "_id": 0,
        "__alias_0": "$_id.__alias_0",
        "__alias_1": "$_id.__alias_1",
        "__alias_2": 1
      }
    },
    {
      "$project": {
        "x": "$__alias_0",
        "y": "$__alias_2",
        "color": "$__alias_1",
        "_id": 0
      }
    },
    {
      "$group": {
        "_id": {
          "x": "$x"
        },
        "__grouped_docs": {
          "$push": "$$ROOT"
        }
      }
    },
    {
      "$sort": {
        "_id.x": 1
      }
    },
    {
      "$unwind": "$__grouped_docs"
    },
    {
      "$replaceRoot": {
        "newRoot": "$__grouped_docs"
      }
    },
    {
      "$limit": 5000
    }
  }
]
```

```
[
  {
    "$group": {
      "_id": {
        "__alias_0": "$Year",
        "__alias_1": "$Reporting_Airline"
      },
      "__alias_2": {
        "$sum": "$LateAircraftDelay"
      }
    }
  },
  {
    {
      "$project": {
        "_id": 0,
        "__alias_0": "$_id.__alias_0",
        "__alias_1": "$_id.__alias_1",
        "__alias_2": 1
      }
    },
    {
      "$project": {
        "x": "$__alias_0",
        "y": "$__alias_2",
        "color": "$__alias_1",
        "_id": 0
      }
    },
    {
      "$group": {
        "_id": {
          "x": "$x"
        },
        "__grouped_docs": {
          "$push": "$$ROOT"
        }
      }
    },
    {
      "$sort": {
        "_id.x": 1
      }
    },
    {
      "$unwind": "$__grouped_docs"
    },
    {
      "$replaceRoot": {
        "newRoot": "$__grouped_docs"
      }
    },
    {
      "$limit": 5000
    }
  }
]
```

# V. TABLEAU DE BORD



AIRLINE DASHBOARD

By  
Oumar Kane